

POLICY RESEARCH WORKING PAPER

5190

Paying Primary Health Care Centers for Performance in Rwanda

Paulin Basinga
Paul J. Gertler
Agnes Binagwaho
Agnes L.B. Soucat
Jennifer R. Sturdy
Christel M.J. Vermeersch

The World Bank
Human Development Network
Chief Economist's Office
&
Africa Region
Health, Nutrition & Population Unit
January 2010



Abstract

Paying for performance (P4P) provides financial incentives for providers to increase the use and quality of care. P4P can affect health care by providing incentives for providers to put more effort into specific activities, and by increasing the amount of resources available to finance the delivery of services. This paper evaluates the impact of P4P on the use and quality of prenatal, institutional delivery, and child preventive care using data produced from a prospective quasi-experimental evaluation nested into the national rollout of P4P in Rwanda. Treatment facilities were enrolled in the P4P scheme in 2006 and comparison facilities were enrolled two years later. The incentive effect is isolated from the resource effect by increasing comparison facilities' input-based budgets by the average P4P payments to

the treatment facilities. The data were collected from 166 facilities and a random sample of 2158 households. P4P had a large and significant positive impact on institutional deliveries and preventive care visits by young children, and improved quality of prenatal care. The authors find no effect on the number of prenatal care visits or on immunization rates. P4P had the greatest effect on those services that had the highest payment rates and needed the lowest provider effort. P4P financial performance incentives can improve both the use of and the quality of health services. Because the analysis isolates the incentive effect from the resource effect in P4P, the results indicate that an equal amount of financial resources without the incentives would not have achieved the same gain in outcomes.

This paper—a joint product of the Chief Economist's Office, Human Development Network, and of the Health, Nutrition & Population Unit, Africa Region—is part of a larger effort to evaluate the impact of innovative programs to improve human development outcomes. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at cvermeersch@worldbank.org, asoucat@worldbank.org, and/or gertler@haas.berkeley.edu.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Paying Primary Health Care Centers for Performance in Rwanda

Paulin Basinga ^{*}

National University of Rwanda School of Public Health

Paul J. Gertler ^{* +}

University of California, Berkeley

Agnes Binagwaho

Ministry of Health, Rwanda

Agnes L.B. Soucat

The World Bank

Jennifer R. Sturdy

The World Bank

Christel M.J. Vermeersch

The World Bank

^{*}Lead Authors, other authors listed alphabetically.

⁺ Corresponding Author:

Paul J. Gertler

Li Ka Shing Professor

Haas School of Business, MC#1900

University of California

Berkeley, CA 94705 USA

Gertler@haas.berkeley.edu

INTRODUCTION

Despite the fact that official development assistance for health has increased dramatically over the last few years, there has been limited improvement in health outcomes in low and middle income countries.^{1, 2, 3} One reason is that health worker productivity and morale is low, often plagued by absenteeism.⁴ One promising intervention to improve worker productivity is pay for performance (P4P),⁵⁻⁹ which provides incentives in the form of bonuses to providers for improvements in utilization and quality of care indicators. P4P can affect health care in two ways; first by incentivizing providers to put more effort into specific activities, and second by increasing the amount of resources available to finance the delivery of services. Despite the promise of P4P, there is little rigorous evidence on its impact in middle and low-income countries, and none that separates out the effect of incentives from increased resources.¹⁰

This study examines the impact of P4P on maternal and child health services in Rwanda, one of the poorest countries in the world.¹¹ Rwanda has made remarkable progress in improving health since the 1994 genocide.¹² This study uses data produced from a prospective quasi-experimental evaluation design nested into the P4P program rollout. Not only is this among the first rigorous evaluations of P4P in a low-income setting, it is the first to isolate the impact of P4P incentives from the associated increase in resources.

METHODS

In 2005, based on positive field reports from pilot NGO P4P schemes, the Government of Rwanda (GoR) decided to implement a national P4P scheme to supplement primary health care centers' input-based budgets with bonus payments based on the quantity and quality of key services.¹³ The scheme pays for 14 maternal and child healthcare output indicators (Box 1).

Some of these indicators are reasons for the visit, such as prenatal care or delivery, while others are services provided during the visit, such as giving the tetanus vaccine during prenatal care. The Ministry of Health (MoH) defined these indicators and payments based on health priorities, available budget and the previous NGO experience.^{13, 14}

Facilities report their monthly indicators to the district P4P steering committee responsible for authorizing payment. For the referral indicators, the facility must also submit verification from the hospital that the referral was appropriate and the referred patient was treated. The committee verifies the reports by sending auditors to the facilities on a quarterly basis on an unannounced randomly chosen day. The auditors review the utilization registry and facility records to verify the data reported is the same as the data recorded in facility records. A comparison of facility records with face-to-face interviews of a random sample of patients reported very little false reporting.¹⁶

The facility's overall quality enters the payment formula as a multiplicative factor that raises or lowers the payment for all outputs. The quality index is bounded between zero and one. If the facility meets all of the quality criteria, then the index equals one and the facility receives full payment for the Box 1 services. However, if the facility is deficient in some of the quality criteria, then all of the payments are discounted. For example, if the facility only scores 0.80 on the quality index, then it only receives 80 percent of the payment for the P4P services.

The facility's overall quality is measured as an index of both structural and process measures of quality of care for various types of services (Box 2).¹⁵ Structural measures are the extent to which the facility has the equipment, drugs, supplies and personnel necessary to deliver a specific service, while process measures capture the clinical content of care provided for

specific services. The structural and process indicators are based on the Rwandan clinical practice guidelines.^{13, 14}

The quality indicators in Box 2 are assessed through the regular monitoring system, in which district hospitals monitor and supervise the quality of health centers in their districts. Every quarter, supervisors from the district hospital visit each facility on an unannounced randomly chosen day, and assess quality indicators through direct observation and review of patient records. At the end of the visit, they discuss their findings with the facility's personnel, and provide recommendations to improve quality of services. In P4P districts, information from the visit is used to compute each facility's overall quality score.

P4P payments go directly to facilities and are used at each facility's discretion. In the sample of 80 treatment facilities in the study, the P4P payments increased average overall expenditures by 22 percent. On average, facilities allocated 77 percent of the P4P funds to increase personnel compensation, amounting to a 38 percent increase in staff salaries.

Experimental Design

The evaluation design took advantage of the phased implementation of the program at the district level over a 24-month period. Administrative districts with pre-existing NGO P4P schemes were excluded from the sampling frame. The remaining districts were then grouped into 8 pairs based on similar characteristics for rainfall, population density, and predominant livelihoods using data from the 2002 Census. One side of each pair was then randomly assigned into the treatment group and the other into the comparison group.

Just before implementation of the baseline survey, the administrative district boundaries were redrawn in the context of a decentralization effort.¹⁷ As a result, some of the experimental

areas were combined with areas that already had the NGO P4P schemes. Because P4P could not be “removed” from health facilities that were already implementing the system, the GoR required that all facilities within those new districts be reassigned to treatment. This led the evaluation team to switch the assignment of treatment and comparison for two of the eight pairs. As a result, 166 of Rwanda’s 401 primary care facilities were included in the study.

The facilities in the treatment group started receiving P4P in 2006, while the facilities in the comparison group continued with input-based financing for an additional 24 months. P4P was implemented in treatment districts over a 5-month period, yielding a minimum 18-month exposure period. Since a primary objective of the evaluation was to isolate the P4P incentive effect separately from the effect of an increase in financial resources, it was necessary to hold the level of resources constant across treatment and comparison facilities. To accomplish this, comparison facilities’ input-based budgets were increased by the average amount of P4P payments to treatment facilities on a quarterly basis during the 24-month treatment window.

We conducted a baseline survey and a follow-up survey 24 months later of the 166 facilities, as well as a random sample of 13 households living in each facility’s catchment area. For the household sample, we first sampled 13 zones (each with approximately 15-20 households) from each facility’s official list of zones in their catchment area. We then physically listed all households in the sampled zones and randomly selected one household with at least one child under 6 years old from each zone. This sample size is large enough to estimate program effects of at least 30% in the number of prenatal care visits and the probability of an institutional delivery with 90% power and a significance level of 0.05 based on data and intra-cluster correlations from the 2005 Rwandan DHS data in rural areas.

Two percent of sampled households refused to participate in the interview. In the follow-up survey, 88 percent of the baseline households were re-interviewed. The rate of attrition from the baseline sample was not statistically different between the treatment and comparison groups (12 percent each). Households that could not be found or interviewed were replaced with randomly selected households from the same zones.

Outcome Measures

Maternal health services: The outcomes include one indicator variable for any prenatal care utilization, one for completing 4+ prenatal care visits, and one for institutional delivery.

Quality of prenatal care: We assess the quality of prenatal care delivered by computing the share of clinical content items that should compose a typical first prenatal consultation, as recommended in the Rwandan clinical practice guidelines for prenatal care, to the actual clinical content items delivered during a prenatal care consultation.¹⁸ This measure has been used extensively in the literature to measure quality.¹⁹⁻²⁵ The items in the measure cover medical history questions, physical examinations, lab tests and follow-up procedures (Box 3). The Cronbach alpha scale of reliability for the 38 item score is 0.78, indicating satisfactory internal consistency. In the impact analysis, we standardize the score by subtracting out the baseline mean and dividing by the baseline standard deviation. The unit of measurement is interpreted as standard deviations in quality.

The prenatal quality score was computed at the individual patient level using two samples. First, enumerators interviewed women who visited the facility as they exited the facility. Second, the same information was collected in the household survey from women who gave birth in the last 18 months and received prenatal care from the facility in whose catchment

area they lived. In the analysis, we combined the exit interview and household survey data for first prenatal care visits to assess the impact of P4P on prenatal care quality, and we control for the observation's source in the analysis with an indicator variable for whether the observation was from the exit interview or household survey.

In addition to the prenatal care process quality index, we measure whether women reported receiving the tetanus vaccine during prenatal care visits, as this enters the payment formula as one of the 14 P4P output indicators and as one of the process quality of care measures in the quality of care index.

Child Preventive Care: We define preventive care as whether the mother reported taking her child for a preventive care visit in the last four weeks. Child preventive services cover immunization, vitamin A, distribution of mosquito nets and child growth monitoring with referral of malnourished children to hospitals for treatment. We analyze preventive visits separately for children 0-23 and 24-59 months old as the younger group is expected to have more visits than the older group.

We also examine the impact of P4P on full immunization, measured by an indicator variable for 12-23 month olds, coded 1 if the child received all vaccines required by the national protocol and 0 otherwise. Vaccination status was assessed based on the vaccination card. Less than 4 percent of the mothers could not produce the child's vaccination card and they were dropped from the analysis.

Statistical Methods

As a result of the reassignment of some districts, we view the evaluation design as quasi-experimental and use difference-in-differences to estimate program impact. This method

compares the change in outcomes in the treatment group to the change in outcomes in the comparison group. By comparing changes, we control for observed and unobserved time-invariant characteristics as well as for time-varying factors that are common to the treatment and comparison groups. The change in the comparison group is an estimate of the true counterfactual – i.e. what would have happened to the treatment group if there had been no intervention.

We estimate a multivariate regression specification of the difference-in-difference model where an individual's outcome is regressed against a dummy variable indicating whether the facility was receiving P4P that year, an individual fixed effect, a year indicator, and a series of individual and household characteristics. We compute robust standard errors clustered at the district by year level to correct for correlation of the error terms across facilities within districts. The models were estimated using Stata version 10.

Covariates

In all of the estimated models, we control for the number of family members, the number of family members under six years old, family assets, the distance of the household from the facility, and whether the family is enrolled in health insurance. Assets are measured as the value of owned houses, durables in the house, farm animals, farm equipment and microenterprise equipment. We also include a variable for if the household owned land. In the prenatal care and delivery analyses, we also control for maternal years of schooling, marital status, whether the partner currently lives in the household, the number of prior pregnancies, and age defined in years. In the prenatal care quality analysis, we also include a variable indicating whether the observation was from the facility exit interview or household survey. In the children's utilization

and outcomes analyses, we include controls for whether the mother and father currently live in the household, their age and years of schooling, the child's age and sex.

Ethical Review

This study was designed and implemented while Gertler was at the World Bank. Although the World Bank does not have an Institutional Review Board, the study team submitted the research protocol to the Rwanda National Ethics Committee, which approved the research design, methodology and methods for informed consent.

RESULTS

We confirm that the evaluation design achieved balance at baseline of observed facility characteristics between the treatment and comparison groups. Table 1 reports the baseline means of facility characteristics in 2006. We find no significant differences between the treatment and comparison groups in terms of log expenditures, allocation of the budget across medical personnel, medical supplies and non-medical purposes, and the numbers of physicians, nurses and other types of personnel.

In the second row of Table 1, we report the mean 2008 log expenditures for treatment and comparison facilities and find no statistically significant difference in the means after the introduction of P4P in the treatment facilities. This confirms that the program compensated the comparison facilities with an increase in their traditional budget equal to the increase in treatment facilities' resources and validates the interpretation of any estimated impacts being caused by in the introduction of P4P incentives, as opposed to an increase in resources.

Maternal Health Services

Our sample consists of women who gave birth in the 18 months prior to the survey: 620 treatments and 670 comparisons. We find no statistical differences in baseline means of the control and outcome variables used in the regression models, except for the share of women who completed 4 prenatal care visits (Table 2a). Although 95 percent of women in the sample obtained some prenatal care, only about 10 percent of women initiated prenatal care in their first trimester and the mean number of visits was less than 3. While about 18 percent of women in treatment areas completed 4 visits, only 11 percent in comparison areas did so. Less than 35 percent of women delivered in a health facility. Finally, while close to 70 percent of women received a tetanus shot, providers executed only 46 percent of the clinical activities recommended in the Rwandan clinical practice guidelines for prenatal care.

Table 3 reports the estimated P4P program impacts on maternal outcomes. First, as indicated by the coefficient on the wave dummy, there were statistically significant increases from 2006 to 2008 in the entire sample for prenatal care, institutional deliveries and quality of prenatal care. There appears to be no impact of P4P on the probability of any prenatal care or on the probability of completing 4+ visits. However, we estimate a statistically significant impact on the probability of institutional delivery of 0.074, with a 95 percent confidence region of (0.004, 0.143), which represents a 21 percent increase from baseline. We also find significant impacts on the quality of prenatal care. We estimate that the impact on the probability of receiving a tetanus vaccine is 0.054 (0.007, 0.265), which is a 7.6 percent increase from the baseline. We also estimate an increase of 0.14 (0.015, 0.265) standard deviations in the standardized prenatal quality score.

Child Health Services

The sample includes children 0-59 months at the time of the survey, and is well balanced between the treatment and comparison groups in terms of baseline control and outcome used in the regression analysis (Table 3). On average across the sample, children 0-23 months old had 0.22 preventive visits in the 4 weeks prior to the survey. Children 24-59 months old living in treatment areas visited the health center 0.08 times in the last 4 weeks on average, while for those living in comparison areas visited 0.14 times, though the difference is not statistically significant at the 5 percent level.

We find large significant impacts of P4P on children's use of preventive health care (Table 4). We estimate that P4P increased the probability that a child 0-23 months visited a health center for preventive care by 0.134 (0.045, 0.224), which is a 64 percent increase over baseline. Similarly, we estimate that P4P increased the probability that a child 24-59 months had a preventive visit by 0.106 (0.050, 0.111), which is a whopping 133 percent increase from the baseline probability for the treatment group. However, we did not find a significant effect on the probability that a child 12-23 months old was fully immunized.

DISCUSSION

We provide evidence that the incentives in the Rwandan P4P program are significantly associated with increased use and quality of a number of critical maternal and child health care services, but not associated with others. These differences are probably related to the structure of the incentives. In general, we see a larger impact on services with higher incentives and for services that are more in the control of the provider and depend less on patients' decisions. For example, prenatal care quality and tetanus vaccination are completely in the provider's control

whereas prenatal care visits depend on the mother deciding to come to the facility. At most, the provider can lobby the mother, but in the end the decision is hers.

Deliveries have the highest unit payment rate at \$4.59. Providers reported they found deliveries to be so lucrative they not only encouraged women to deliver in the facility during prenatal care, but some also commissioned community health workers to conduct outreach in the community to find pregnant women to deliver in the facility.

Similarly, the large increase in preventive child visits is also explained by the higher payment rate. While the payment rate per child preventive visit is low at \$0.18, the rate for finding a malnourished child and referring her for treatment is very high at \$1.83. Given that almost 50 percent of children in Rwanda are stunted and could be referred,¹² approximately half of the child preventive growth monitoring visits yielded \$0.18 each, while the other half yielded \$2.01 each for the growth visit plus the identification and referral of a malnourished child.

One of the strongest monetary payoffs is for prenatal care quality. Specifically, every administration of tetanus vaccine and malaria prophylaxis yields \$0.92, as well as increases the prenatal care quality index score. In addition, providers can use prenatal care visits to lobby women to deliver in the facility, a service for which they receive an additional \$4.59. Finally, improved compliance with prenatal care clinical practice guidelines CPGs raises the facility's overall quality score and thereby the share of the P4P payments actually received.

The lack of impact on prenatal care utilization is also explained by poor financial incentives. The payment rate for the initial prenatal visit is very low at \$0.09. Since over 95 percent of women at baseline made at least one visit, the \$0.09 payment provides little incentive to find the few remaining women who do not use prenatal care. Second, the payment rate for completing 4 visits is only \$0.37. Since women start prenatal care late in Rwanda (5th or 6th

month) and have between 2 to 3 visits on average, it is difficult to complete additional visits in such a short period of time. Providers may not have wanted to expend the enormous effort necessary to get these women to come back for 1 to 2 more visits for only \$0.37.

The fact that we did not find an impact of P4P on child vaccinations might be explained in part by the fact that baseline immunization rates were close to 65% and the government implemented an intensive national vaccination campaign starting in 2006.²⁶ An increase beyond this would have required substantial effort on the part of the providers to enter the community, identify unvaccinated children and provide them with vaccinations. Moreover, completing the immunization program for a child requires a lot of effort as it involves many clinic visits over the course of the year and more complex record keeping.

One possible alternative explanation for our findings is increased monitoring and supervision led to increased provider effort. However, there are a number of reasons to believe that the financial incentives are driving the results. First, monitoring and supervision by the district hospitals is required nationally, independent of the P4P scheme. Second, if the primary pathway was through increased monitoring and supervision, one would expect to see an increase in provider knowledge in the treatment facilities as a result of the feedback during the supervision visits. We find no statistical difference between treatment and comparison groups on provider knowledge on prenatal care. Third, we only see improvement in some indicators and not in others. If the monitoring were effective, then we would see increases in both low priced services and high priced services.

Policy Implications

Our findings illuminate the debate on some of the benefits and shortfalls of P4P. A number of specific P4P lessons also emerge from this study.

1. Higher payments provide high-powered incentives.
2. Incentives have a larger impact for services in which providers have more control, such as prenatal care quality. Therefore, higher incentive payments are warranted not only for services that are more important in terms of leading to better health outcomes, but also where more provider effort is required to improve those services.
3. Programs should consider paying more for verifiable clinical content indicators. Content indicators are closely related to outcomes, are measurable, and are solely within the control of the provider.
4. For services that depend more on patient behavior, such as the decision to seek prenatal care, the program might provide financial incentives directly to the patient rather than the provider. Indeed, there is substantial evidence that conditional cash transfers to families increase preventive care utilization and improve health outcomes.⁴¹⁻⁴⁹ Another feasible alternative is to give community health workers an incentive to identify patients and encourage them to visit clinics.

One of the more important results of this analysis is the effect of P4P on the quality of care delivered. Although health workers may be competent to perform a medical procedure or consultation (i.e. prenatal care), they may not always be willing or motivated to expend the effort to perform all the required components of that procedure.^{27, 28} By conditioning the Rwanda P4P payment on a quality index score, the evidence suggests that the incentive gave providers the motivation to translate their knowledge about prenatal care into better practice. These results are

important primarily because better quality care yields better health outcomes, and in the case of prenatal care this impacts both pregnant women and their children.^{27, 29-32} Indeed it is not enough to get mothers and children into care, it also matters that the care they receive is of high quality.

Limitations

There are a number of limitations to the study. First, the original randomized designed was revised due to the political decentralization process. While we can be confident in the results of the difference-in-differences analysis given that the sample is well balanced on observable characteristics and outcomes at baseline, this is one of the challenges for effectiveness studies in the context of the scale-up of national programs. However, while small-scale efficacy trials are easier to control, they only inform us about what is possible given the best-case scenario in program implementation. Efficiency studies, such as this rigorous evaluation of a national health program, are important for external validity and inform us about what is the likely impact under more realistic conditions. Second, we only measure quality of care for prenatal services. Hence further study is required to see if the incentive effects for prenatal care quality extend to other services.

ACKNOWLEDGEMENTS

We gratefully acknowledge funding from grants from The British Economic and Social Research Council, The Global Development Network, a Japanese PHRD grant, The World Bank's Bank-Netherlands Partnership Program, and The World Bank's Spanish Impact Evaluation Fund. Dr. Basinga also gratefully acknowledges financial support from Global Development Network and the MacNamara Foundation. While the evaluation design was developed in close consultation with the Rwandan P4P project team, all data collection, analysis, interpretation, preparation, and review of this manuscript were independent of any influence by the P4P program staff. We the authors also declare that we have no conflict of interest.

We are also indebted to Claude Sekabaraga, Louis Rusa, Rigobert Mpendwanzi, Kathy Kantengwa, Gyuri Fritsche, and Bruno Meessen for useful comments on the study design, discussions on program design and implementation, and continued support of the study. We also thank Lia Fernald, Anastasia Gage, Paul Hutchinson, Benjamin Loevinsohn, Nancy Mock, and Nancy Padian for valuable comments on early drafts. We are grateful to the Rwandan Ministry of Health, Management Sciences for Health, the Belgian Technical Cooperation, Cordaid, GTZ, Healthnet, USAID, and the World Bank and their staff for their collaboration in the implementation of the P4P rollout plan and supporting the evaluation. For data collection efforts, we thank the National University of Rwanda School of Public Health and CIEDEP. Finally, we gratefully acknowledge Elena Samonte-Hinckley, Rosangela Bando, Rafael Ramirez and Ricardo Cavazos for research assistance. The authors are solely responsible for the content and opinions presented in this paper. The content and opinions in this paper do not necessary represent the view of the funders or of the organizations that employee the authors.

Box 1: Output Indicators and Unit Payments for P4P Formula

OUTPUT INDICATORS		Amount paid per unit (US\$)
Visit and Outreach Indicators: Number of ...		
1	curative care visits	0.18
2	first prenatal care visits	0.09
3	women who completed 4 prenatal care visits	0.37
4	first time family planning visits (new contraceptive users)	1.83
5	one-month contraceptive resupply	0.18
6	deliveries in the facility	4.59
7	child (0 - 59 months) growth monitoring (preventive care) visits	0.18
8	children who completed vaccinations on time	0.92
Content of care indicators: Number of ...		
9	women who received appropriate tetanus vaccine during prenatal care ⁺	0.46
10	women who received 2nd dose of malaria prophylaxis during prenatal care	0.46
11	at risk pregnancies referred to hospital for delivery during prenatal care ⁺⁺	1.83
12	emergency transfers to hospital for obstetric care during delivery ⁺⁺	4.59
13	malnourished children referred for treatment during preventive care visit ⁺⁺	1.83
14	other emergency referrals during curative treatment ⁺⁺	1.83

⁺ Appropriate is defined to any woman who obtains her second, third, fourth or fifth tetanus shot.

⁺⁺ Referrals must be confirmed by hospital that patient was treated and referral was necessary.

Box 2: Services and Weights Used to Construct the Quality Score for P4P Formula

(Web Version ONLY)

		Quality	Share of weight	Share of weight	
	Service	Index	allocated to	allocated to	Means of assessment
		Weight	structural	process	
			components	components	
1	General administration	0.052	1.00	0.00	Direct observation
2	Cleanliness	0.028	1.00	0.00	Direct observation
3	Curative care	0.170	0.23	0.77	Medical record review
4	Delivery	0.130	0.40	0.60	Medical record review
5	Prenatal care	0.126	0.12	0.88	Direct observation
6	Family planning	0.114	0.22	0.78	Medical record review
7	Immunization	0.070	0.40	0.60	Direct observation
8	Growth monitoring	0.052	0.15	0.85	Direct observation
9	HIV services	0.090	1.00	0.00	Direct observation
10	Tuberculosis services	0.028	0.28	0.72	Direct observation
11	Laboratory Services	0.030	1.00	0.00	Direct observation
12	Pharmacy management	0.060	1.00	0.00	Direct observation
13	Financial management	0.050	1.00	0.00	Direct observation
	Total	1.000			

Box 3: Items Included in Rwandan Clinical Practice Guidelines for Prenatal Care
(Web Version ONLY)

Area	Items included
Pregnancy history:	Number of prior pregnancies, living children, and miscarriages, bleeding during previous labor, how the last child was delivered, and last child's birth weight.
Pregnancy history:	Number of prior pregnancies, living children, and miscarriages, bleeding during previous labor, how the last child was delivered, and last child's birth weight.
Gynecological history:	STIs including HIV, pap smear, contraceptive use, last menstrual date, and related health problems.
Medical history:	High blood pressure, diabetes, contraceptive use, heart disease, disease, malaria, goiter, and tobacco and alcohol use.
Obstetric symptoms:	Contractions, vaginal bleeding, weight loss/gain, nausea, vomiting, and current medications.
Physical examination:	Body height, body weight, check vital signs (blood pressure, temperature, respiratory), palpitate abdomen, listen to fetal heartbeat, check for edema and measure uterus.
Laboratory tests	Hemoglobin (anemia), diabetes, urine protein, platelet count, HIV and STIs (syphilis /gonorrhea).
Prevention/case management:	Advice about nutrition, tetanus vaccine, iron/folic acid supplementation, advice about danger signs for emergency help, HIV voluntary counseling/test, and complete prenatal card.

Table 1: Health Facility Baseline (2006) Characteristics

	Treatment		Comparison		Difference	P-Value*
	Mean	Std dev	Mean	Std dev		
Expenditures and Budget Shares						
Log Total Expenditures (2006)	15.81	(1.04)	15.61	(1.01)	0.200	0.418
Log Total Expenditures (2008)	16.91	(0.71)	16.99	(1.08)	-0.083	0.568
Medical Personnel Budget Share	0.46	(0.23)	0.49	(0.26)	-0.031	0.555
Medical Supplies Budget Share	0.22	(0.19)	0.20	(0.19)	0.013	0.705
Non-medical Budget Share	0.32	(0.25)	0.30	(0.22)	0.018	0.726
Staffing						
Medical Doctors	0.05	(0.23)	0.05	(0.27)	0.003	0.940
Nurses	6.31	(6.90)	5.48	(3.30)	0.826	0.409
Other Clinical Staff	4.13	(3.09)	4.47	(4.05)	-0.335	0.554
Non-clinical Staff	5.25	(3.56)	5.33	(5.09)	-0.076	0.901
Observations	80		86			

Note: All variables, except Log Expenditures 2008, are measured at baseline (2006).

*P-values are for cluster-adjusted t-test (continuous variables).

Table 2: Maternal Sample Baseline (2006) Characteristics

	Treatment		Comparison		Difference	p-value*
	Mean	Std Dev	Mean	Std Dev		
Maternal characteristics						
Age	30.89	(7.05)	31.22	(6.85)	-0.33	0.594
Primary education or more (=1)	0.42	(0.03)	0.35	(0.04)	-0.01	0.889
Currently married/union (=1)	0.94	(0.23)	0.91	(0.29)	0.04	0.214
Partner present (=1)	0.98	(0.14)	0.97	(0.17)	0.01	0.325
Number of pregnancies (Parity)	4.32	(2.46)	4.33	(2.43)	-0.01	0.969
Number of living children	3.39	(1.93)	3.51	(2.62)	-0.12	0.490
Health insurance (=1)	0.55	(0.50)	0.52	(0.50)	0.04	0.591
Household characteristics						
Number of household members	5.15	(1.70)	5.40	(1.94)	-0.25	0.145
Own land (=1)	0.93	(0.25)	0.87	(0.33)	0.06	0.126
Value of Assets (household, animals, farm and enterprise equipment)						
<i>Quartile 1</i> (=1)	0.22	(0.42)	0.22	(0.42)	0.00	0.994
<i>Quartile 2</i> (=1)	0.25	(0.43)	0.25	(0.43)	0.00	0.885
<i>Quartile 3</i> (=1)	0.27	(0.45)	0.28	(0.45)	-0.01	0.885
<i>Quartile 4</i> (=1)	0.25	(0.43)	0.24	(0.43)	0.01	0.772
Maternal care utilization						
Any prenatal care (=1)	0.95	(0.21)	0.96	(0.20)	-0.01	0.773
Started in first trimester	0.11	(0.32)	0.09	(0.29)	0.02	0.544
4 or more prenatal care visits (=1)	0.18	(0.38)	0.11	(0.31)	0.07	0.036
Number of prenatal care visits	2.76	(0.84)	2.62	(0.80)	0.14	0.180
Delivery in facility (=1)	0.35	(0.48)	0.36	(0.48)	-0.01	0.801
Quality of prenatal care						
Total Quality Score	0.46	(0.02)	0.47	(0.02)	-0.01	0.6661
Standardized Total Quality Score	-0.13	(0.06)	-0.10	(0.06)	-0.03	0.7348
Tetanus vaccine (=1)	0.71	(0.45)	0.67	(0.47)	0.04	0.309

* P-values are for cluster-adjusted t-test or chi-squared tests of difference.

Table 3: Children 0-5 years old Baseline (2006) Characteristics

	Treatment		Comparison		Difference	P-Value*
Variable	Mean	Std Dev	Mean	Std Dev		
Child characteristics						
Age (in months)	30.45	(0.58)	29.95	(0.58)	0.50	0.51
Female (=1)	0.50	(0.01)	0.49	(0.01)	0.01	0.59
Parental characteristics						
Mother present (=1)	0.97	(0.01)	0.96	(0.01)	0.01	0.13
Mother's age	31.62	(0.33)	32.08	(0.32)	-0.45	0.38
Mother's years of education	5.57	(0.19)	5.61	(0.20)	-0.03	0.99
Father present (=1)	0.87	(0.02)	0.90	(0.02)	-0.03	0.30
Father's age	36.10	(0.39)	36.58	(0.40)	-0.48	0.33
Father's years of education	6.06	(0.31)	6.27	(0.34)	-0.20	0.65
Household characteristics						
Number of household members	5.06	(0.09)	5.21	(0.09)	-0.15	0.26
Number of members 0-5 years old	2.18	(0.03)	2.21	(0.03)	-0.03	0.44
Own land (=1)	0.87	(0.02)	0.92	(0.02)	-0.05	0.15
Value of Assets (household, animals, farm and enterprise equipment)						
Quartile 1 (=1)	0.22	(0.03)	0.23	(0.02)	-0.01	0.77
Quartile 2 (=1)	0.28	(0.03)	0.25	(0.02)	0.03	0.27
Quartile 3 (=1)	0.26	(0.04)	0.28	(0.03)	-0.02	0.71
Quartile 4 (=1)	0.23	(0.02)	0.24	(0.02)	-0.01	0.71
Child preventive medical care						
Visit by child 0-23 months in last 4 weeks (=1)	0.21	(0.02)	0.24	(0.03)	-0.03	0.56
Visit by child 24-59 months in last 4 weeks (=1)	0.08	(0.02)	0.14	(0.02)	-0.05	0.11
12-23 month old is fully immunized (=1)	0.62	(0.06)	0.67	(0.06)	-0.05	0.52

* P-values are for cluster-adjusted t-test or chi-squared tests of difference.

Table 4: Estimated Impact of P4P on Maternal and Child Health Care Services

	N	β	(95% C.I.)	P-Value	% Δ^+
Maternal Care Utilization*					
Any prenatal care (=1)	2223	0.002	(-0.022 - 0.025)	0.894	0.2%
Made 4 or more prenatal care visits	2223	0.010	(-0.063 - 0.083)	0.790	5.6%
Institutional delivery (=1)	2108	0.074	(0.006 - 0.142)	0.033	21.1%
Quality of Prenatal Care**					
Standardized Quality Score	3683	0.14	(0.015 - 0.265)	0.030	NA
Tetanus vaccine during prenatal visit (=1)	2810	0.054	(0.007 - 0.100)	0.024	7.6%
Child Preventive Visit in last 4 weeks***					
Visit by child age 0-23 months (=1)	1973	0.134	(0.045 - 0.224)	0.004	63.8%
Visit by child age 24-59 months (=1)	3645	0.106	(0.050 - 0.161)	0.000	132.5%
Child 12-23 months is fully immunized (=1)	732	-0.065	(-0.178 - 0.047)	0.248	-10.5%

* The β is the estimated treatment effect controlling for a year dummy, facility fixed effects, individual-level characteristics (age, education, number of children, civil status, presence of partner, health insurance) and household characteristics (number of household members, value of assets, land ownership and distance from the facility). Standard errors were adjusted for clustering at the health facility-year level.

** The β is the estimated treatment effect controlling for provider-level characteristics (age, gender, competency score), patient-level characteristics (age, education, civil status, insurance enrollment). Standard errors were adjusted for clustering at the health facility-year level. The number of observations in the tetanus model is less than in the quality score model because tetanus is only given to women with 5 pregnancies or less.

*** The β is the estimated treatment effect controlling for individual-level characteristics (age, gender, insurance enrollment), parental-level characteristics (mother/father present, age, education) and household characteristics (number of household members, value of assets, land ownership and distance from the facility). Standard errors were adjusted for clustering at the health facility level.

⁺ The % $\Delta = (\beta / \text{Baseline Mean}) \times 100$, where the baseline mean of the dependent variable is the 2006 mean of the treatment group from Tables 2 and 3.

References

- (1) Horton R. Countdown to 2015: a report card on maternal, newborn, and child survival. *Lancet* 2008 Apr 12;371(9620):1217-1219.
- (2) World Bank. Global Monitoring Report. 2006.
- (3) Gottret P, Schieber G. *Health Financing Revisited*. : The World Bank; 2008.
- (4) Chaudhury N, Hammer J, Kremer M, Muralidharan K, Rogers FH. Missing in action: teacher and health worker absence in developing countries. *J.Econ.Perspect.* 2006 Winter;20(1):91-116.
- (5) Levine R, Eichler R. Performance Incentives for Global Health. 2009.
- (6) Brenzel L, Bredenkamp C, Naimoli J, Batson A, Skolnik R, Measham AR. Taking stock: World Bank experience with results-based financing (RBF) for health. 2009.
- (7) Meessen B, Kashala JP, Musango L. Output-based payment to boost staff productivity in public health centres: contracting in Kabutare district, Rwanda. *Bull.World Health Organ.* 2007 Feb;85(2):108-115.
- (8) Meessen B, Musango L, Kashala JP, Lemlin J. Reviewing institutions of rural health centres: the Performance Initiative in Butare, Rwanda. *Trop.Med.Int.Health* 2006 Aug;11(8):1303-1317.
- (9) Soeters R, Habineza C, Peerenboom PB. Performance-based financing and changing the district health system: experience from Rwanda. *Bull.World Health Organ.* 2006 Nov;84(11):884-889.
- (10) Eldridge C, Palmer N. Performance-based payment: some reflections on the discourse, evidence and unanswered questions. *Health Policy Plan.* 2009 May;24(3):160-166.
- (11) World Bank. World Development Indicators Database. 2008.
- (12) Rwanda 2005: results from the Demographic and Health Survey. *Stud.Fam.Plann.* 2008 Jun;39(2):147-152.
- (13) Ministry of Health. Republic of Rwanda. Guide for Performance Based Financing. Training module for actors involved in the implementation of the PBF program. 2006.
- (14) Rwanda Ministry of Health. Reference Module for Family Planning training. For use of trainers, supervisors and providers at the health facility level. March 2008.

- (15) Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988 Sep 23-30;260(12):1743-1748.
- (16) HDP Rwanda. Report on Community Counter Verification in Nyamasheke, Nyanza, Nyaruguru and Rulindo districts. December 2008
- (17) MINALOC (Ministry of Local Government). Making decentralized service delivery work in Rwanda: putting the people at the center of service provision. Kigali: Government of Rwanda, Policy Note. 2006.
- (18) Ministry of Health Contractual Approach Unit, Rwanda. Quarterly quality assessment grid for health centers. 2008.
- (19) Barber S. Does the quality of prenatal care matter in promoting skilled institutional delivery? A study in rural Mexico. *Matern.Child Health J.* 2006 Sep;10(5):419-425.
- (20) Barber SL, Bertozzi SM, Gertler PJ. Variations in prenatal care quality for the rural poor in Mexico. *Health.Aff.(Millwood)* 2007 May-Jun;26(3):w310-23.
- (21) Barber SL, Gertler PJ. Empowering women to obtain high quality care: evidence from an evaluation of Mexico's conditional cash transfer programme. *Health Policy Plan.* 2009 Jan;24(1):18-25.
- (22) Peabody JW, Gertler PJ. Are clinical criteria just proxies for socioeconomic status? A study of low birth weight in Jamaica. *J.Epidemiol.Community Health* 1997 Feb;51(1):90-95.
- (23) Peabody JW, Luck J, Glassman P, Jain S, Hansen J, Spell M, et al. Measuring the quality of physician practice by using clinical vignettes: a prospective validation study. *Ann.Intern.Med.* 2004 Nov 16;141(10):771-780.
- (24) Peabody JW, Rahman O, Fox K, Gertler P. Quality of care in public and private primary health care facilities: structural comparisons in Jamaica. *Bull.Pan Am.Health Organ.* 1994 Jun;28(2):122-141.
- (25) Das J, Hammer J. Which Doctor? Combining vignettes and item response to measure doctor quality. May 2004;Policy Research Working Paper 3301.
- (26) Ministry of Health, Rwanda. Annual report 2007. 2008.
- (27) Kogan MD, Alexander GR, Kotelchuck M, Nagey DA. Relation of the content of prenatal care to the risk of low birth weight. Maternal reports of health behavior advice and initial prenatal care procedures. *JAMA* 1994 May 4;271(17):1340-1345.
- (28) Leonard KL, Masatu MC. The use of direct clinician observation and vignettes for health services quality evaluation in developing countries. *Soc.Sci.Med.* 2005 Nov;61(9):1944-1951.

- (29) Villar J, Ba'aqeel H, Piaggio G, Lumbiganon P, Miguel Belizan J, Farnot U, et al. WHO antenatal care randomised trial for the evaluation of a new model of routine antenatal care. *Lancet* 2001 May 19;357(9268):1551-1564.
- (30) Villar J, Bergsjö P. Scientific basis for the content of routine antenatal care. I. Philosophy, recent studies, and power to eliminate or alleviate adverse maternal outcomes. *Acta Obstet.Gynecol.Scand.* 1997 Jan;76(1):1-14.
- (31) Carroli G, Rooney C, Villar J. How effective is antenatal care in preventing maternal mortality and serious morbidity? An overview of the evidence. *Paediatr.Perinat.Epidemiol.* 2001 Jan;15 Suppl 1:1-42.
- (32) Carroli G, Villar J, Piaggio G, Khan-Neelofur D, Gulmezoglu M, Mugford M, et al. WHO systematic review of randomised controlled trials of routine antenatal care. *Lancet* 2001 May 19;357(9268):1565-1570.
- (33) Scaling up innovation in services delivery of Primary Health Care: Case of Rwanda. ; Jan 30 - Feb 1, 2008; ; 2008.
- (34) World Bank, Ministry of Health. Country status report on health and poverty in Rwanda. 2009.
- (35) Bell J, Hussein J, Jentsch B, Scotland G, Bullough C, Graham W. Improving skilled attendance at delivery: a preliminary report of the SAFE strategy development tool. *Birth* 2003 12//;30(4):227-234.
- (36) Bulatao RA, Ross JA. Which health services reduce maternal mortality? Evidence from ratings of maternal health services. *Trop.Med.Int.Health* 2003 08//;8(8):710-721.
- (37) de Bernis L, Sherratt DR, AbouZahr C, Van Lerberghe W. Skilled attendants for pregnancy, childbirth and postnatal care. *Br.Med.Bull.* 2003;67:39-57.
- (38) Improve access to quality maternal health services. ; 97; Colombo: Technical consultation on safe motherhood; 1997.
- (39) Kayongo M, Butera J, Mboninyibuka D, Nyiransabimana B, Ntezimana A, Mukangamuje V. Improving availability of EmOC services in Rwanda--CARE's experiences and lessons learned at Kabgayi Referral Hospital. *Int.J.Gynaecol.Obstet.* 2006 Mar;92(3):291-298.
- (40) Campbell OM, Graham WJ, Lancet Maternal Survival Series steering group. Strategies for reducing maternal mortality: getting on with what works. *Lancet* 2006 Oct 7;368(9543):1284-1299.
- (41) Fiszbein A, Schady N. Conditional cash transfers: reducing present and future poverty 2009.

- (42) Fernald LC, Gertler PJ, Neufeld LM. Role of cash in conditional cash transfer programmes for child health, growth, and development: an analysis of Mexico's Oportunidades. *Lancet* 2008 Mar 8;371(9615):828-837.
- (43) Barber SL, Gertler PJ. The impact of Mexico's conditional cash transfer programme, *Oportunidades*, on birthweight. *Tropical Medicine & International Health* 2008;13(11):1405-1414.
- (44) Gertler PJ. Final report: the impact of Progesa on health. November 2000.
- (45) Gertler PJ. Do conditional cash transfers improve child health? Evidence from PROGRESA's control randomized experiment. *American Economic Review* 2004;94:336-341.
- (46) Lagarde M, Haines A, Palmer N. Conditional cash transfers for improving uptake of health interventions in low- and middle-income countries: a systematic review. *JAMA* 2007 Oct 24;298(16):1900-1910.
- (47) Morris SS, Flores R, Olinto P, Medina JM. Monetary incentives in primary health care and effects on use and coverage of preventive health care interventions in rural Honduras: cluster randomised trial. *Lancet* 2004 Dec 4-10;364(9450):2030-2037.
- (48) Rivera JA, Sotres-Alvarez D, Habicht JP, Shamah T, Villalpando S. Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children: a randomized effectiveness study. *JAMA* 2004 Jun 2;291(21):2563-2570.
- (49) Barham T. Providing a healthier start to life: the impact of conditional cash transfer programs on Infant Mortality. January 2005.